Applied Vegetation Science 11: 499-508, 2008
doi: 10.3170/2008-7-18560, published online 4 June 2008

Mapping mountain vegetation using species distribution modeling, image-based texture analysis, and object-based classification

Dobrowski, Solomon Z.[1*]; Safford, Hugh D.[3]; Cheng, Yen Ben[2] & Ustin, Susan L.[2]

[1]*Department of Forest Management, College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA;* [2]*Center for Spatial Technologies and Remote Sensing, One Shields Ave, The Barn, University of California, Davis, CA 95616, USA;* [3]*USDA-Forest Service Pacific Southwest Region, 1323 Club Drive, Vallejo, CA 94592, USA;* [*]*Corresponding author; E-mail solomon.dobrowski@cfc.umt.edu*

Abstract

Objective: The objective of this study was to map vegetation composition across a 24 000 ha watershed.

Location: The study was conducted on the western slope of the Sierra Nevada mountain range of California, USA.

Methods: Automated image segmentation was used to delineate image objects representing vegetation patches of similar physiognomy and structure. Image objects were classified using a decision tree and data sources extracted from individual species distribution models, Landsat spectral data, and life form cover estimates derived from aerial image-based texture variables.

Results: A total of 12 plant communities were mapped with an overall accuracy of 75% and a $\varkappa$-value of 0.69. Species distribution model inputs improved map accuracy by approximately 15% over maps derived solely from image data. Automated mapping of existing vegetation distributions, based solely on predictive distribution model results, proved to be more accurate than mapping based on Landsat data, and equivalent in accuracy to mapping based on all image data sources.

Conclusions: Results highlight the importance of terrain, edaphic, and bioclimatic variables when mapping vegetation communities in complex terrain. Mapping errors stemmed from the lack of spectral discernability between vegetation classes, and the inability to account for the confounding effects of land use history and disturbance within a static distribution modeling framework.

Keywords: Decision tree; GAM; Image segmentation; Sierra Nevada; Topographic convergence index; Vegetation mapping.

Nomenclature: Hickman (ed.) 1993.

Abbreviations: DOQQ = Digital ortho-photo quarter quads; DT = Decision tree; GAM = General additive modeling; SDM = Species Distribution Modeling; TCI = Topographic convergence index.

Introduction

Image classification and predictive distribution modeling (referred to as species distribution modeling, gradient modeling, niche modeling, among others) are common approaches to mapping vegetation. Both approaches have limitations, for instance, mapping based on the classification of spectral data is often hindered by the lack of spectral discernability between vegetation types (Dirnböck et al. 2003; Treitz et al. 1992), whereas predictive distribution modeling characterizes potential rather than actual vegetation distributions (Austin 2002; Guisan & Zimmerman 2000). When used in combination, both image analysis and predictive modeling can be complementary. Terrain variables have long been used in image classification to improve map accuracies (J. Franklin 1995). Similarly, image variables have been used in predictive distribution modeling (Dirnböck et al. 2003; Lees & Ritman 1991) and are more recently referred to as functional gradients (Muller 1998). A challenge to automated mapping of vegetation, is devising a methodology that leverages the strengths of both predictive modeling and image-based approaches, as well as divides finite resources between tasks associated with each.

Approaches to incorporating environmental variables into existing vegetation mapping efforts include but are not exclusive to: (1) the direct use of environment and image variables in a nominal classifier, (2) constrained ordination techniques using inventory data, image and environmental variables (e.g. Ohmann & Gregory 2002; Dirnböck et al. 2003), and (3) Probabilistic or consensus theoretic approaches to combining image analysis and gradient modeling results (e.g. Strahler 1980; Dobrowski et al. 2006). The use of environmental variables in these and others studies has been shown to improve map accuracies given that these types of data are related to biophysical factors that affect the distribution of species (J. Franklin 1995; Richards et al. 1982; Strahler 1980).

Mapping vegetation at high spatial resolution presents unique challenges. Variability of ground targets increase as image resolution becomes finer often resulting in speckled classification results (Martin & Howarth 1989). Further, pixel-based image analysis approaches are limited due to their inability to incorporate contextual or topological information into the classification problem (Debeir et al. 2002; Gong & Howarth 1992; Karathanassi et al. 2000). Despite this, high resolution image data provides unique opportunities for mapping vegetation structure. As image resolution becomes finer than the scale of individual trees and patches of vegetation (i.e. H-resolution image model, Strahler et al. 1986), image texture variables increasingly become correlated to vegetation structure and physiognomy (S. Franklin et al. 2000; Wulder 1998; Wulder et al. 2004).

Automated image segmentation is increasingly being used in vegetation mapping applications (Blaschke et al. 2000; Zhang & Luo 2000). Object-based classification groups spectrally similar pixels together (segmentation) to form image objects and then classifies these objects based on attributes including spectral, textural, and topological features. These techniques have been applied to identify anthropogenic features, burned areas, and individual tree crowns (Benz et al. 2004; Gougeon 1995; Mitri & Gitas 2004) as well as vegetation patches in montane environments (Ryherd & Woodcock 1996; Woodcock & Harward 1992), rangeland communities (Laliberte et al. 2005), chaparral and woodland communities (Shandley et al. 1996), and alpine environments (Dirnböck et al. 2003).

Study objectives

The primary objective of this study was to develop a vector-based map of dominant plant communities for a watershed on the western slope of the Sierra Nevada, California, USA. In contrast to regional mapping efforts (e.g. S. Franklin et al. 2000; Cohen et al. 2001), our intent was to develop a map that has utility at the local watershed scale and provides finely resolved, stand specific information on vegetation composition and structure. A requirement was that the map meet the Federal Geographic Data Committee (FGDC) vegetation classification standard, a physiognomic-floristic hierarchy (Jennings et al. 2003). A secondary objective was to assess the efficacy of common methods and information sources used in vegetation mapping. To these ends, we provide a case study in which we quantify the contribution of individual data sources and methodologies towards improving map accuracy.

Methods

Overview

The overall mapping approach involved three parallel methodologies (Fig. 1) including life form cover estimation from image texture analysis, species distribution modeling (SDM) as a means to incorporate ecological 'context' into the mapping problem, and automated image segmentation to provide a consistent support for extracting data from multiple data sources for use in categorical vegetation type classification.

Study site

Our study site is the watershed of the North Fork of the Middle Fork of the American River (NFMF), located on the western slope of the Sierra Nevada, California, USA (lat. 39°03'47", long. 120”40'40"). The study area is ca. 24 000 ha in size and spans elevations from 320 to 2190 m. The climate of the study area is Mediterranean with cool wet winters and hot dry summers. At the lower elevations, the primary geomorphic features are steep inner-canyon gorges of the NFMF American River and associated tributaries. These environs have extreme topography with many areas having slope gradients greater than 60 degrees. The upper elevations of the study area are comprised of colluvial hillslopes capped by broad volcanic flows. On this physical template are heterogeneous vegetation types including mixed chaparral, constituents of the mixed hardwood conifer group, the Sierra mixed conifer group, and true firs as you move from low to high elevation.

The study area is almost entirely in public holding and is located on the southwestern edge of the Tahoe National Forest. It has a long history of resource extraction including hydraulic mining and timber extraction. In this respect, the study area is representative of many national forest lands of similar elevation on the west slope of the Sierra Nevada. A large portion of the upper reaches of the study area show effects of even- and uneven-aged forest management practices.

In situ measurements

In situ data for the project was collected to support both SDM and image object based classification. These data included 338 fixed area plots ('inventory plots') used in SDM and ca. 700 GPS points ('reference points') used in developing training and validation data for the image object-based classification. Inventory plots were used to provide species presence and absence data for SDM. Prior to locating inventory plots, we stratified the landscape based on three environmental surfaces including elevation,

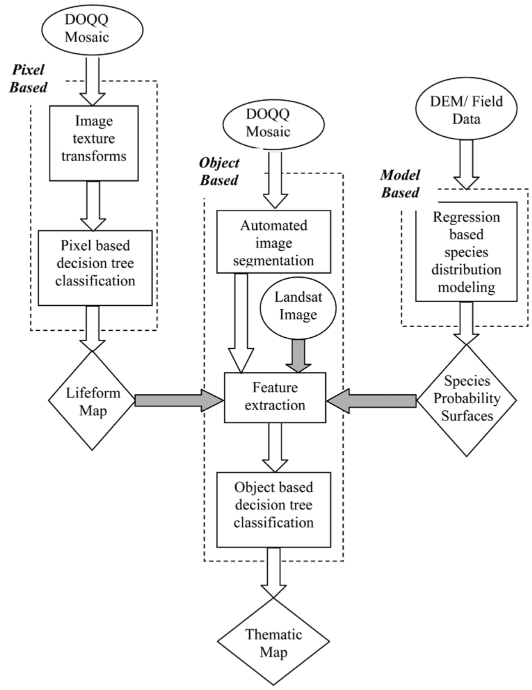


Fig. 1. Work flow chart of mapping procedures.

geology, and topographic relative moisture index (Parker 1982). These surfaces were combined to locate a representative sample of plots within unique eco-regions. Data collection in each circular 800 m^2 inventory plot followed protocols outlined by Winthers et al. (2003).

Over 700 reference points were collected using GPS. Our objective was to collect a large number of training and validation reference points to aid in typing image objects to vegetation classes using color aerial photos. Reference points were taken directly within vegetation stands by field crew members and typed using the CALVEG system (Parker & Matayas 1979) using keys and descriptions developed by the USDA-Forest Service Pacific Southwest Region.

Species Distribution Modeling (SDM)

Presence-absence data from the inventory plots were used for modeling the distribution of 23 tree and shrub species. We utilized general additive modeling (GAM), a semi-parametric form of regression for SDM. GAM was implemented using the GRASP (Lehmann et al. 2002a) toolset in S-plus (Insightful Inc.). Data screening and model fitting procedures followed methods described by Dobrowski et al. (2006). We produced raster surfaces (30-m resolution) characterizing the probability of species presence for each of the 23 species. Examples of these surfaces are provided in App. 1.

Predictor variables

Eleven predictor variables were used in the SDM analysis. Six of these variables were produced from a 30-m USGS digital elevation model including elevation, slope, potential annual direct radiation (RAD - McCune & Keon 2002), topographic convergence index (TCI - Wolock & McCabe 1995), transformed aspect (TRAN – Beers et al. 1966), and relative slope position (RSPS – Wilds 1996). TCI quantifies available soil moisture as a function of topographic position and slope. TRAN was calculated as COS(45-aspect) and is an indirect measure of heat loading. RSPS identifies local topographic minima and maxima to identify valley bottoms and ridge tops. The remaining five explanatory variables included two factor variables (six bedrock geology classes (Loyd 1995) and fire frequency; for the latter see: http://frap.cdf.ca.gov/data/ frapgismaps /select.asp) as well as three climate variables including minimum annual temperature, maximum annual temperature, and average precipitation acquired from PRISM (www.ocs.orst.edu/prism/) at 4-km resolution.

Model assessment

We qualitatively assessed models by comparing response curves to known autecological species characteristics (e.g. Lehmann et al. 2002b) . A quantitative assessment of the modeling results was conducted using a five-fold cross-validation procedure. Cross-predicted Pearson correlation coefficients and the area under the curve statistic - AUC (Fieldings & Bell 1997) were used to quantify model quality for the binomial model. Both the cross correlation and the AUC statistics were calculated from the average of five ‘leave one out’ iterations.

Image analysis and Object-based classification

Image data

Digital ortho-photo quarter quads (DOQQ) were used for high resolution (1 m) panchromatic imagery; 21 DOQQ images, acquired in the summer months of 1998, were mosaiced over the study area. We texture transformed the DOQQ mosaic using a grey level co-occurrence matrix (GLCM), five texture algorithms (Haralick et al. 1973), and three window sizes using ENVI image analysis software (Research Systems Inc., Boulder, CO). Additionally, Landsat ETM+ imagery from June of 2002 was acquired in order to provide multi-spectral data at a 30 m resolution. A summary of all primary image data, pre and post processing analysis steps, and derived raster datasets is provided in Table 1.

Table 1. Summary of primary raster data sources, pre- and post-processing steps, derived raster data sources, and uses within the study. See Methods for further detail.

Primary raster data	Pre-processing steps	Post-processing steps	Derived raster data	Spatial resolution (m)	Use
DOQQ		Resampling		3	Image segmentation
	Mosaicing,	GLCM texture transform at 3,5,7 m windows	Mean, contrast, Entropy, ASM, correlation	1	Life form classification
Landsat ETM+	Topographic shade correction	Spectral transform	Landsat spectral bands (1,2,3,4,5,7), Normalized difference vegetation index NDVI = (band4 -band3)/ (band4 + band3)	30	Image object-based classification
USGS DEM		See Methods	Elevation, slope, aspect, Potential radiation (RAD), Topographic convergence index (TCI), Transformed aspect (TRAN), Relative slope position (RSPS)	30	Species distribution modeling

Image segmentation

Automated image segmentation was conducted on the DOQQ mosaic using a region-based segmentation algorithm (eCognition, Definiens Inc.) with parameters iteratively determined using subsets of imagery, a qualitative assessment by an analyst, and feedback from resource managers. Ca. 10 000 vector polygons ('image objects') were created that represent vegetation patches.

Life form classification and cover estimates

Four life forms were identified in the DOQQ imagery including hardwood, conifer, shrub, and herbaceous/ barren classes. Training and validation data for the classifier was collected by digitizing a minimum of 100 polygons in the DOQQ mosaic for each life form class. Image texture bands were used as attributes with a decision tree classifier (See 5.0; Rulequest Inc.) based on recursive partitioning (Breiman et al. 1984). The classifier was trained with 70% of the training set with the remainder used for assessing accuracy. A boosting algorithm was used with 10 trials to improve the classifier performance. The trained classifier was used to develop a nominal map of life form (1 m resolution) from which calculations of percent cover by life form per image object were made.

Due to height and relief displacement in the DOQQ images, cover estimates for tree life forms were increasingly overestimated as one moved from the DOQQ image center towards the tile boundaries. To correct this bias we developed and applied the following regression equation to the conifer, hardwood, and shrub cover estimates:

$$c_u = \beta_0 + \beta_0 c_b + \beta_2 d \qquad (1)$$

where c_u is the unbiased corrected cover estimate, c_b is the biased texture image derived cover estimate, d is the distance from each image object centroid to the nearest DOQQ tile center point, and β's are derived regression coefficients. Unbiased cover estimates ($c\mathrm{B}_{u\mathrm{B}}$) to fit this equation were developed from digitized aerial photos and point-intercept sampling (see digital appendix 2 for further detail).

Image object-based vegetation classification

We identified 12 CALVEG 'alliances' to use in automated mapping. For each class, we identified a minimum of 30 image objects for use in classifier training and accuracy assessment. This was accomplished through manual photo interpretation of color aerial photos verified by typed reference points. Table 2 lists the CALVEG classes used in the analysis and the number of image object samples identified for each class.

For each image object, 78 attributes were extracted from three primary data sources: 1. Landsat data (mean, max, min, standard deviation for each of 6 spectral bands) 2. Percent life form cover estimates for each of the 4 life form types (conifer, hardwood, shrub, herbaceous/barren) 3. Mean and standard deviation of GAM derived probability estimates for each of 23 species.

We assessed the utility of each data source used in the analysis by training a decision tree (DT) with different combinations of the available data: (1) All data sources, (2) Landsat spectral data and life form cover data, (3) SDM probability surfaces and Landsat spectral data, (4) life form cover estimates only, (5) Landsat spectral data only, and (6) SDM probability surfaces only.

Table 2. Summary of map classes and per-class sample size for the image object-based classification used in the study. CALVEG is a regional map product of dominant vegetation types developed by the USDA Forest Service.

Vegetation class	CALVEG Code	Dominant species	Image object sample size
Greenleaf Manzanita	CG	*Arctostaphylos patula*	30
Huckleberry Oak	CH	*Quercus vaccinifolia*	31
Whiteleaf Manzanita	CW	*Arctostaphylos viscida*	30
Upper Montane Mixed Chaparral	CX	*Arctostaphylos patula, Arctostaphylos viscida, Ceanothus cordulatus, Ceanothus integerrimus*	33
Douglas Fir - Pine	DP	*Pseudotsuga menziesii, Pinus ponderosa*	33
Mixed Conifer - Fir	MF	*Abies concolor, Abies magnifica, Calocedrus decurrens, Pinus jeffreyi, Pinus contorta, Pinus lambertiana*	32
Mixed Conifer - Pine	MP	*Abies concolor, Calocedrus decurrens, Pinus ponderosa, Pinus lambertiana, Pseudotsuga menziesii*	73
Ponderosa Pine	PP	*Pinus ponderosa*	35
Canyon Live Oak	QC	*Quercus chrysolepis*	36
Black Oak	QK	*Quercus kelloggii*	30
Red Fir	RF	*Abies magnifica*	30
White Fir	WF	*Abies concolor*	30

For each combination described above, a DT (SEE5.0) was trained on 70% of the data (293 image objects) and overall accuracy was assessed with the remainder. A boosting algorithm with 10 trials was used to improve classifier results. Class level accuracies were determined separately due to the small sample size for the test set on a class by class basis. An independent decision tree was constructed with the entire data set and class level accuracy was determined using a 10-fold cross-validation procedure.

Results

Species Distribution Modeling

When constructing SDMs, we found that in all instances, the three climatic variables were highly correlated with elevation ($r > 0.80$). Consequently, we dropped them from the analysis due to their coarse resolution and to avoid model instability due to collinearity.

We highlight general trends in SDM construction: (1) Elevation was the primary indirect gradient identified in terms of model contribution for most species;

Table 3. Summary of GAM model quality by species. AUC represents the 'area under the curve' statistic and is scaled from 0.5 (poor model fit) to 1.0 (excellent model fit). Results are the average of a 5-fold cross validation.

	Species	Sample size	Cross correlation (r)	AUC
TREES	*Abies concolor*	252	0.51	0.79
	Abies magnifica	94	0.66	0.89
	Acer macrophyllum	305	0.50	0.85
	Alnus rhombifolia	216	0.29	0.79
	Calocedrus decurrens	299	0.58	0.83
	Pinus attenuata	192	0.22	0.72
	Pinus jeffreyii	62	0.54	0.74
	Pinus lambertiana	283	0.42	0.74
	Pinus monticola	63	0.50	0.80
	Pinus ponderosa	312	0.46	0.76
	Psuedotsuga menziesii	321	0.53	0.79
	Quercus chrysolepis	210	0.53	0.79
	Quercus kelloggii	289	0.56	0.81
SHRUBS	*Arctostaphylos nevadensis*	116	0.77	0.95
	Arctostaphylos patula	144	0.63	0.90
	Arctostaphylos viscida	188	0.64	0.89
	Ceanothus cordulatus	64	0.43	0.87
	Ceanothus integerrimus	196	0.45	0.77
	Chrysolepis sempervirens	46	0.36	0.86
	Quercus chrysolepis	48	0.30	0.79
	Quercus vaccinifolia	122	0.77	0.95
	Ribes spec.	144	0.27	0.69
	Salix spec.	37	0.12	0.61

Table 4. Summary of life form classification accuracies derived from the classification of DOQQ image texture variables.

	Producer (%)	User (%)
Hardwood	65	70
Conifer	91	87
Shrub	52	75
Herbaceous/barren	99	98

(2) Predictor variables related to water availability (TCI, RSPS) and evaporative demand (transformed aspect, RAD) were significant predictors for most species; (3) Fire frequency (a measure of disturbance) was found to be a significant predictor for a number of fire adapted, chaparral, and early successional species (e.g. *Pinus attenuata, Quercus kelloggii,* and *Ceanothus* spec.).

Quantitative metrics of model quality are summarized in Table 3. The results from these statistics show a wide range of model accuracies with most models showing satisfactory to good fits (AUC values from 0.7 to 0.9).

Image segmentation

Ca. 10 000 image objects varying in size from 0.058 ha to 27 ha (mean of 2.4 ha, median of 1.77 ha), were created using image segmentation. Height displacement of trees near the DOQQ image boundaries occasionally resulted in erroneous polygon linework across image boundaries. In some cases, polygon boundaries were manually dissolved. Fig. 2 shows a subset of image objects within the study area.

Life form classification and cover estimates

Results from the life form classification are summarized in Table 4. Image texture proved to be an effective attribute for distinguishing between life forms. Conifers were readily identified (user accuracy = 87%) whereas the primary confusion occurred between hardwood and shrub types (results not shown).

As expected, there was systematic bias in the DOQQ

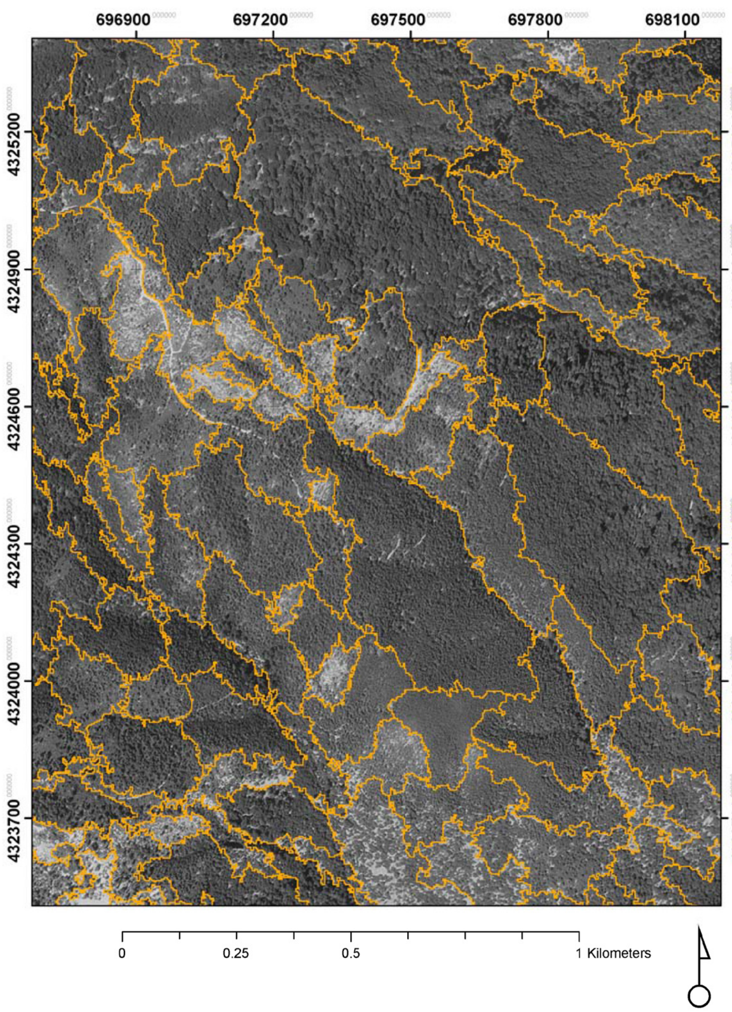


Fig. 2. Subset of digital orthophoto-quarter quad (DOQQ) mosaic showing image objects derived from automated segmentation.

Table 5. Summary of accuracy statistics for object-based classification based on multiple data sources. Overall accuracy (OA) was calculated with 30% of the data set held out for validation. Class level producer and user accuracies were calculated with an independent decision tree using a 10-fold cross validation procedure.

	Accuracy (%)											
	All attributes OA = 75%		Landsat and % cover by life form OA = 62%		Landsat and SDM* OA = 71%		% cover by life form OA = 42%		Landsat data OA = 42%		SDM* OA = 62%	
Map Class #	producer	user	producer	user	producer	user	producer	user	producer	user	producer	user
CG	80	83	63	70	70	68	27	36	57	61	60	58
CH	90	82	81	69	77	77	55	47	68	53	71	59
CW	71	71	36	43	68	63	11	20	39	39	39	50
CX	79	74	67	59	73	69	48	32	33	39	42	45
DP	76	76	58	59	79	79	27	29	55	60	76	78
MF	64	54	58	55	64	55	16	25	55	44	49	43
MP	51	67	31	37	51	69	56	41	23	33	49	59
PP	78	85	58	58	78	82	14	15	58	55	75	69
QC	70	75	73	63	67	74	47	52	70	68	30	36
QK	77	79	63	63	77	72	3	6	60	53	70	64
RF	43	59	43	48	40	57	60	47	23	39	53	57
WF	65	59	55	55	65	72	43	46	60	60	70	58
mean	70	72	57	57	67	70	32	31	50	50	57	56

* Species distribution model (SDM) probability surfaces;
\# See Table 2 for class definitions.

derived tree cover estimates that was positively related to distance from the DOQQ tile center points. The correction procedure improved cover estimates across the entire range of cover values. Cross validated RMSE estimates of life form cover are as follows: conifer (8.3%), hardwood (9.6%), shrub (14%). Given that the reference cover values used in this analysis were derived from an independent data source (aerial photos), the RMSE estimates presented incorporate both errors in the nominal life form classification and errors due to bias correction. App. 3 provides an example of cover correction for conifers.

Image object-based vegetation classification

Accuracy results for all combinations of information sources used in the DT classifier are presented in Table 5. Plant communities were mapped with an overall accuracy of 75% and a κ-value of 0.69. Overall accuracies calculated using a hold-out test dataset were similar to or higher than those calculated on a class by class basis using cross validation. When using all the information sources as attributes, the average producer and user accuracies were 70% and 72% respectively. When using the image derived data only (Landsat spectral data and % cover by life form), average producer and user accuracies were 57%, suggesting that the modeled species distribution data improved the classification results by roughly 15%. Interestingly, map accuracies calculated with SDM data were greater than those calculated using Landsat data (producer and user accuracies of 57% vs. 50%) and are similar to those calculated when using *all* image data sources (57%). The exclusion of percent life form cover estimates (Landsat and SDM attributes only) resulted in map accuracies that were 2% to 3% less than those calculated using all attributes, suggesting that life form cover estimates carried little additional information relevant to floristic mapping.

Discussion

Map accuracy

Mapping efforts have varied objectives, extents, spatial and thematic resolution, and approaches to accuracy assessment. The direct comparison of map accuracies should be done with caution. With this in mind, we note that our accuracy results are consistent with, or improved over previous mapping efforts within the region that employed CALVEG. Compared to results from J. Franklin et al. (2001), life form classification accuracies from this study were greater (70-98% versus 62-79%) while vegetation type accuracies were similar (54-85% versus 54-84%). One limitation of our approach is that we did not directly incorporate land-use history information or collect field data within recently harvested or disturbed sites. Consequently, our vegetation type accuracies are likely to be overstated when taking into account the full range of land-use types found within the region. Our cover estimates were more accurate than previous approaches using Landsat data. For example, we calculated an RMSE of 8.3% and 10% for continuous conifer and hardwood cover estimates respectively. This is in contrast to 23% RMSE for conifer cover noted in Cohen et al. (2001) and 28-69% and 27% calculated for 10% conifer and hardwood cover classes respectively in J. Franklin et al. (2001). We note that the previous studies were mapping efforts conducted at regional scales, whereas the current study is more local in its domain.

Image data and its utility

Image objects contained vegetation of similar physiognomy and structure but varied in floristic composition. Given the use of DOQQs in this study, it is clear that the segmentation process did not delineate objects based on spectral differences but instead relied on image texture differences associated with vegetation structure. Our results are in agreement with previous studies that show that spectral variability is likely to play a secondary role to image texture in driving the region-growing segmentation process (Ryherd & Woodcock 1996; Greenberg et al. 2006).

Image objects were poorly typed using spectral data. This was due to categorical vegetation classes with inherent mixed composition (e.g. mixed conifer types) and poor spectral discernability between types using Landsat data. Previous work has shown the limitations of using Landsat data in vegetation mapping of forests (S. Franklin 2001). We do not elaborate on this further other than to emphasize that the utility of Landsat data seems to lie principally in mapping land cover types, physiognomy of existing vegetation, and some forms of vegetation structure (Cohen et al. 2001; Lees & Ritman 1991; Ohmann & Gregory 2002) as opposed to site-specific floristic composition in mixed species stands.

Considerable effort was dedicated to estimating life form cover using the DOQQ imagery. Unfortunately, the inclusion of life form cover in the classifier did little to improve nominal vegetation type accuracies. In this case, knowing that a vegetation patch had 80% versus 90% conifer cover did little to improve our ability to type the stand as compared to just knowing that it was conifer dominated - information that was likely inherent in the coarser scaled Landsat data (Cohen et al. 2001). Additionally, cover estimation could have been greatly simplified through the use of high spatial resolution satellite imagery acquired in a nadir direction (e.g. IKONOS or Quickbird) to avoid relief displacement bias. Despite this, the DOQQs proved to have utility in developing continuous cover estimates at the patch scale - data not readily available through coarser resolution data-sources or through photo-interpretation.

Single species distribution modeling in multispecies mapping

In contrast to direct gradient modeling of categorical vegetation types, we chose to use single species SDMs. Our approach was to predict the probability of presence of individual plant species and treat these predictions as 'composition' data within a categorical vegetation type classification. In the present study, we did not cluster or aggregate single species predictions (e.g. J. Franklin 2002), but instead used a decision tree to 'weight' the SDM and image-derived information sources on a class-by-class basis given categorical vegetation types defined *a priori*. Decision trees are well suited for this type of problem given that a DT can effectively optimize a large number of weighting combinations (*sensu* Benediktsson & Sveinsson 2003) and it can handle non-linear interactions (De'Ath & Fabricius 2000).

The use of an SDM implies the use of an ecological model (Austin 2002) that assumes that vegetation communities are comprised of species assembled along biophysical gradients (the familiar 'continuum' concept of vegetation community ecology (Whittaker 1951)). If we assume that *species* respond to environmental gradients, then a relevant issue is the extent to which categorical vegetation types represent mutually exclusive groups of species. If vegetation types are narrowly defined, then environment-response functions will be more precise. If vegetation types have significant overlap in species composition, then derived environment-response functions will lack specificity. Many of the vegetation types encountered within our study region have mixed assemblages of species and substantial overlap in species composition between types (e.g. mixed conifer). In this context, single species SDMs minimize the number of questionable assumptions about the reality and stability of categorical vegetation types and their response to environmental gradients.

Single species models often demonstrate improved predictions over multi-species model predictions for the same species (Guisan et al. 1999). Many multi-species ordination approaches assume a Gaussian species-environment response. A unimodal and symmetric response may or may not be expected when examining an entire species distribution (Austin 2002), however, for local mapping efforts such as this, truncated response curves are to be expected. Non- or semi-parametric single species modeling approaches (e.g. GAM) allow for the development of complex and truncated response functions. Additionally, the single species modeling approach used here allows the map to retain continuous species-level information along with a nominal vegetation type.

Combining predictive distribution modeling, image texture analysis, and object-based classification was a profitable approach to mapping vegetation in an area of complex terrain. Our findings suggest that incorporating environmental data into the mapping problem (through SDM or other gradient modeling approaches) is critical in mapping efforts where thematic classes are based on floristic units and where spectral discernment using imagery remains a challenge.

Acknowledgements. We would like to thank John Babin and Vivian Kee with the Tahoe National Forest for their support with this work. Also, thanks to our field crew Missy Voight, Ben Addlestone, Jennifer Riddell, and Dylan Burge. Lastly, we thank Dr. Janet Franklin, Dr. Gretchen Moisen, and Dr. Ken Brewer for their valuable comments on the manuscript. This research was funded by the USDA Forest Service, Pacific Southwest Region.

References

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecological Modeling* 157: 101-118.

Beers, T.W., Press, P.E. & Wensel, L.C. 1966. Aspect transformation in site productivity research. *Journal of Forestry* 64: 691-692.

Benediktsson, J.A. & Sveinsson, J.R. 2003. Multisource remote sensing data classification based on consensus and pruning. *IEE Transactions on Geoscience and Remote Sensing* 41: 932-936.

Blaschke, T., Lang, S., Lorup, E., Strobl, J. & Zeil, P. 2000. Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. In: *Environmental information for planning, politics and the public*. Metropolis-Verlag, Marburg, DE.

Breiman, L., Friedman, J.H., Olshend, R.A. & Stone, C.J. 1984. *Classification and regression trees*. Chapman and Hall, NY, US.

Cohen, W.B., Maiersperger, T.K., Spies, T.A. & Oetter, D.R. 2001. Modeling forest cover attributes as continuous variables in a regional context with Thematic Mapper data. *International Journal of Remote Sensing* 22: 2279-2310.

De'Ath, G. & Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178-3192.

Debeir, O., Van den Steen, I., Latinne, P., Van Ham, P. & Wolff, E. 2002. Textural and contextual land-cover classification using single and multiple classifier systems. *Photogrammetric Engineering and Remote Sensing* 68: 597-605.

Dirnböck, T., Dullinger, S., Gottfried, M., Ginzler, C. & Grabherr, G. 2003. Mapping alpine vegetation based on image analysis, topographic variables and canonical correspondence analysis. *Applied Vegetation Science* 6: 85-96.

Dobrowski, S.Z., Greenberg, J.A., Ramirez, C.M. & Ustin, S.L. 2006. Improving image derived vegetation maps with regression based distribution modeling. *Ecological Modeling* 192: 126-142.

Fieldings, A.H. & Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.

Franklin, J. 1995. Predictive vegetation mapping: Geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474-499.

Franklin, J. 2002. Enhancing a regional vegetation map with predictive models of dominant plant species in chaparral. *Applied Vegetation Science* 5: 135-146.

Franklin, J., Simons, D., Beardsley, D., Rogan, J.M. & Gordon, H. 2001. Evaluating errors in a digital vegetation map with forest inventory data and accuracy assessment using fuzzy sets. *Transactions in Geographic Information Systems* 5: 285-304.

Franklin, S.E. 2001. *Remote sensing for sustainable forest management*. CRC Press, Boca Raton, FL, US.

Franklin, S.E., Hall, R.J., Moskal, L.M., Maudie, A.J. & Lavigne, M.B. 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing*. 21: 61-79.

Gong, P. & Howarth, P.J. 1992. Frequency-based contextual classification and gray-level vector reduction for land-use identification. *Photogrammetric Engineering and Remote Sensing*. : 423-437.

Gougeon, F.A. 1995. A crown following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*. 21: 274-284

Greenberg, J.A., Dobrowski, S.Z., Ramirez, C.M., Tuil, J.L. & Ustin, S.L. 2006. A bottom-up approach to vegetation mapping of the Lake Tahoe Basin using hyperspatial image analysis. *Photogrammetric Engineering and Remote Sensing* 72: 581-589.

Guisan, A. & Zimmerman, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modeling*. 135: 147-186.

Guisan, A., Weiss, S. & Weiss, A. 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology* 143: 107-122.

Haralick, R.M., Shanmugam, K. & Dinstein, I. 1973. Texture feature for image classification. *IEE Transactions on Systems, Man, and Cybernetics* 2: 610-621.

Hickman, J.C. (ed.) 1993. *The Jepson Manual*. University of California Press, Berkeley, CA, US.

Jennings, M., Loucks, O., Glenn-Lewin, D., Peet, R., Faber-Langendoen, D. et al. 2003. *Guidelines for describing associations and alliances of the U.S. national vegetation classification*. Vegetation Classification Panel, Ecological Society of America, Washington, DC, US.

Karathanassi, V., Iossifidis, C. & Rokos, D. 2000. A texture based classification method for classifying built areas according to their density. *International Journal of Remote Sensing* 21: 1807-1823.

Laliberte, A.S., Rango, A. & Fredrickson, E.L. 2005. *Classification of arid rangelands using an object-oriented and multi-scale approach with quickbird imagery.* Geospatial goes global: from your neighborhoods to the whole planet.

ASPRS. Baltimore, March 7-11.

Lees, B.G. & Ritman, K. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management* 15: 823-831.

Lehmann, A., Overton, J. & Leathwick, J. 2002a. GRASP: generalized regression analysis and spatial prediction. *Ecological Modeling* 157: 189-207.

Lehmann, A., Leathwick, J. & Overton, J. 2002b. Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation* 11: 2217-2238.

Loyd, R.C. 1995. *Geologic map of Placer County, California. 1:100,000 scale*. California Division of Mines and Geology, Sacramento, CA, US.

Martin, L.R.G. & Howarth, P.J. 1989. Change-detection accuracy assessment using spot multispectral imagery of the rural urban fringe. *Remote Sensing of Environment* 30: 55-66.

McCune, B. & Keon, D. 2002. Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science* 13: 603-606.

Mitri, G.H. & Gitas, I.Z. 2004. A semi-automated object-oriented model for burned area mapping in the Mediterranean region using Landsat-TM imagery. *International Journal of Wildland Fire* 13: 367-376.

Muller, F. 1998. Gradients in ecological systems. *Ecological Modeling* 108: 3-21.

Ohmann, J.L. & Gregory, M.J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. *Canadian Journal of Forest Research* 32: 725-741.

Parker, A. J. 1982. The topographic relative moisture index: An approach to soil moisture assessment in mountain terrain. *Physical Geography* 3: 160-168.

Parker, I. & Matayas, W. 1979. *CALVEG: a classification of Californian vegetation*. Region 5 Ecology Group, USDA Forest Service. San Francisco, CA, US.

Richards, J.A., Landgrebe, D.A. & Swain, P.H. 1982. A means for utilizing ancillary information in multi spectral classification. *Remote Sensing of Environment* 12: 463-478.

Ryherd, S. & Woodcock, C.E. 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering and Remote Sensing* 62: 181-194.

Shandley, J., Franklin, J. & White, T. 1996. Testing the Woodcock-Harward image segmentation algorithm in an area of southern California chaparral and woodland vegetation. *International Journal of Remote Sensing* 17: 983-1004.

Strahler, A.H. 1980. The use of prior probabilities in maximum-likelihood classification of remotely sensed data. *Remote Sensing of Environment* 10: 135-163.

Strahler, A.H., Woodcock, C.E. & Smith, J.A. 1986. On the nature of models in remote sensing. *Remote Sensing of Environment* 20: 121-139.

Treitz, P. M., Howarth, P.J. & Suffling, R.C. 1992. Application of detailed ground information to vegetation mapping with high spatial resolution digital imagery. *Remote Sensing of Environment* 42: 65-82.

Whittaker, R.H. 1951. A criticism of the plant association and climatic concepts. *Northwest Science* 25: 17-31.

Wilds, S.P. 1996. *Gradient analysis of the distribution of flowering dogwood and dogwood anthracnose in western Great Smoky Mountains National Park*. M.Sc. Thesis, University of North Carolina. Chapel Hill, NC, US.

Winthers, E., Fallon, D., Haglund, J., Demeo, T. et al. 2003. *Terrestrial Ecological Unit Inventory Technical Guide*. USDA Forest Service, Washington, DC, US.

Wolock, D.M. & McCabe, J. 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research* 31: 1315-1324.

Woodcock, C.E. & Harward, V.J. 1992. Nested hierarchical scene models and image segmentation. *International Journal of Remote Sensing* 13: 3167-3187.

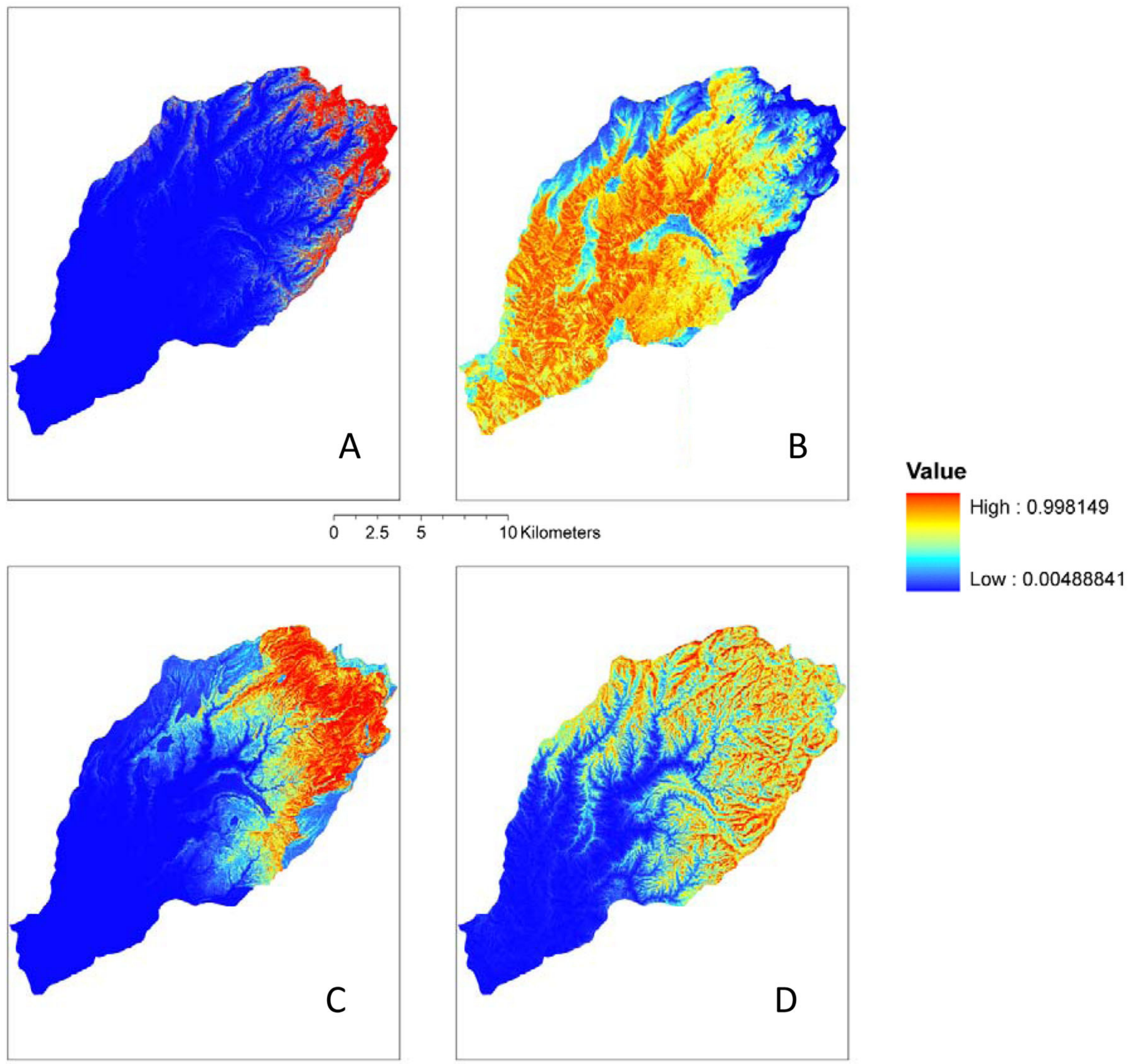
Wulder, M.A. 1998. Optical remote sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography* 22: 449-476.

Wulder, M.A., Hall, R.J., Coops, N.C. & Franklin, S.E. 2004. High spatial resolution remotely sensed data for ecosystem characterization. *Bioscience* 54: 511-521.

Zhang, Y.J. & Luo, H. 2000. Optimal selection of segmentation algorithms based on performance evaluation. *Optical Engineering* 39: 1450-1455.

Received 11 September 2007;
Accepted 8 April 2008;
Co-ordinating Editor: J.L. Ohmann.

*For App. 1-3, see below (online version)
also available at JVS/AVS Electronic Archives;
www.opuluspress.se/*

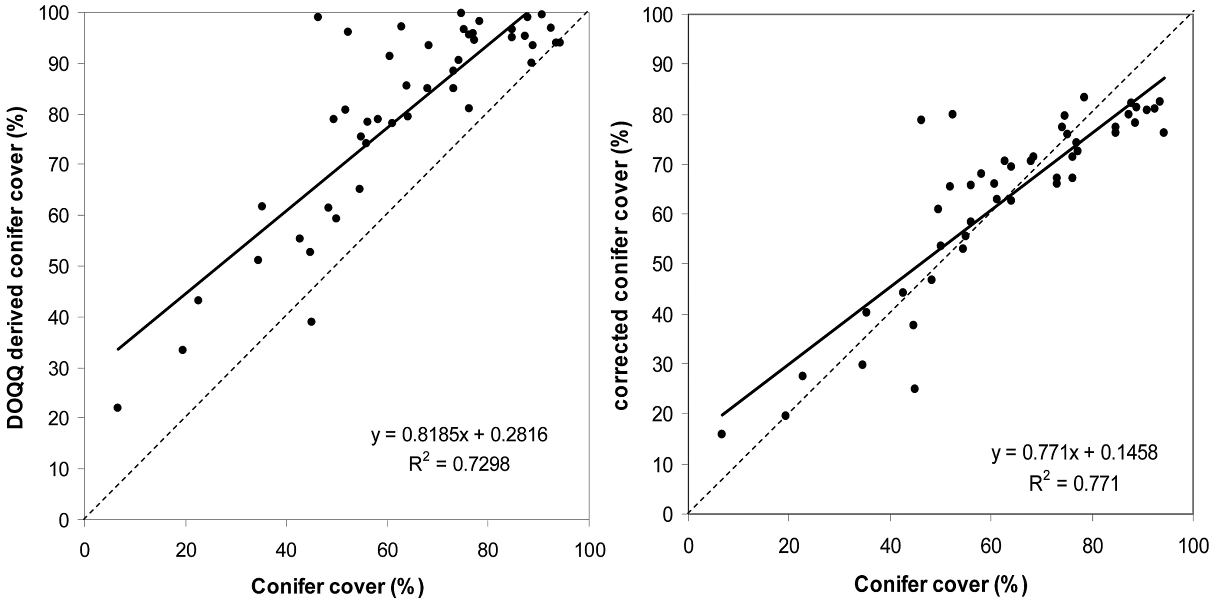


App. 1. Probability of species presence for two tree and shrub species within the study area. **A.** *Abies magnifica*; **B.** *Pseudotsuga menziesii*; **C.** *Quercus vaccinifolia*; **D.** *Arctostaphylos patula*.

App. 1-3. Internet supplement to: Dobrowski, S.Z.; Safford, H.D.; Cheng, Y.B. & Ustin, S.L. 2008.
Mapping mountain vegetation using species distribution modeling, image-based texture analysis, and object-based classification
Applied Vegetation Science 11: 499-508; doi: 10.3170/2008-7-18560

OPULUSPRESS

App. 2. Methods for unbiased life form cover estimation.
A total of 10 aerial photos were digitally scanned and georeferenced to the DOQQ imagery. Polygon boundaries from image segmentation of DOQQ imagery were used to identify sample units in the digitized aerial photos. Only polygons located near the center points of the aerial photos were considered for training data in order to avoid the same height and relief displacement issues found in the DOQQ imagery. Forty-eight image polygons were selected and extracted from the digital aerial photos.
Each of the image polygons were manually classified using SamplePoint software (http://www.ars.usda.gov/services/software/software.htm), a manual image-analysis program designed to facilitate vegetation cover measurements using point intercept sampling. Cover by life form for each image polygon was estimated using a 225 point sample grid. Equation 1 was fit for the conifer, hardwood, and shrub life forms and goodness of fit (root means square error - RMSE) was determined using a five-fold cross-validation procedure. The cover correction was subsequently applied to all image objects in the study area. For a given polygon, total cover of all life forms was constrained to 100% by normalizing all cover estimates by the sum of the corrected estimates.



App. 3. Example of the effect of a bias correction procedure on image object conifer cover estimates. Height displacement in DOQQ image tiles resulted in biased conifer cover estimates. The left panel shows a comparison of biased versus independently assessed conifer cover estimates. The right panel shows the same comparison after a bias correction was applied. See Methods for further details.

App. 1-3. Internet supplement to: Dobrowski, S.Z.; Safford, H.D.; Cheng, Y.B. & Ustin, S.L. 2008. Mapping mountain vegetation using species distribution modeling, image-based texture analysis, and object-based classification
Applied Vegetation Science 11: 499-508; doi: 10.3170/2008-7-18560